04-07-08

EQ044117087US

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF APPEALS
AND INTERFERENCES

Patent Application

Inventors:	**Eric T. Bax** **Charles C. Fowlkes** **Louis Cisnero, Jr.**
Case No.:	**iSpheres 1**
Serial No.:	**10/352,720**
Filing Date:	**January 28, 2003**
Title:	**Normalization of Speech Accent**

Group Art Unit: **2626**
Examiner: **Justin W. Rider**

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

SIR:

FEE FOR FILING A BRIEF IN SUPPORT OF APPEAL

Please charge the amount of $510 to Avaya Inc. Deposit Account No. 501602 to cover the fee for filing a brief in support of an appeal under 37 CFR 1.17(c).

Triplicate copies of this letter are enclosed. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **Deposit Account No. 501602** as required to correct the error.

Respectfully submitted,

Eric T. Bax
Charles C. Fowlkes
Louis Cisnero, Jr.

By David Volejnicek
David Volejnicek
Corporate Counsel
Reg. No. 29355
303-538-4154

Date: 04 April 2008

Avaya Inc.
Docket Administrator
307 Middletown-Lincroft Road, Room 1N-391
Lincroft, NJ 07738

"Express Mail" mailing label No.

I hereby certify that this Application is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Commissioner for Patents, PO Box 1450, Alexandria, VA 22313-1450.

On 04 April 2008	Tina Wilson	
Date of Deposit	(Type name of person mailing paper)	(Signature of person mailing paper)

Serial No. 10/352,720

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

Patent Application

Inventors:	**Eric T. Bax** **Charles C. Fowlkes** **Louis Cisnero, Jr.**
Case No.:	**iSpheres 1**
Serial No.:	**09/728,689** Group Art Unit: **2178**
Filing Date:	**December 1, 2000**
Examiner:	**Gregory J. Vaughn**
Title:	**Technique for Extracting Data from Structured Document**

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

APPLICANTS'/APPELLANTS' APPEAL BRIEF

SIR:

Applicants/Appellants hereby appeal to the Board of Patent Appeals and Interferences in response to the Notice of Panel Decision from Pre-Appeal Brief Review mailed on March 10, 2008. The fee set forth in 37 CFR §41.20(b) has been previously submitted in connection with the Request for Pre-Appeal Brief Request for Review. Although Applicants/Appellants believe that no additional fees are due, authorization is hereby given to charge any necessary fees to Deposit Account No. 501602.

A single copy of this Brief is being submitted pursuant to MPEP §1205.02.

04/07/2008 RFEKABU1 00000024 501602 10352720
01 FC:1402 510.00 DA

TABLE OF CONTENTS

REAL PARTY IN INTEREST

The real party in interest is Avaya Inc, the assignee of the above-identified application, as evidenced by the assignment recorded in the US Patent and Trademark Office on Reel 018131, Frame 0415.

RELATED APPEALS AND INTERFERENCES

None.

STATUS OF CLAIMS

Claims canceled: 1-42

Claims withdrawn from consideration, but not canceled: None

Claims pending: 43-46

Claims allowed: None

Claims rejected: 43-46

Claims objected to: None

Claims appealed: 43-46

STATUS OF AMENDMENTS

No amendments were filed subsequently to the notice of final rejection. A Response to Final Office Action that was filed on 29 November 2007 and that contains only Remarks/Arguments was entered. A Pre-Appeal Brief Request for Review that was filed on 11 January 2008 was entered. A Notice of Panel Decision from Pre-Appeal Brief Review was mailed on 10 March 2008, directing applicants to proceed to the Board of Patent Appeals and Interferences.

SUMMARY OF CLAIMED SUBJECT MATTER

Traditional techniques of extracting information about a subject from data such as a web page or a text file are typically based on knowledge of the structure used to arrange data within each specific web site or page. The structure is commonly referred to as the syntax. (Specification at page 1, lines 21-23.) These traditional techniques are limited because they can only gather attribute values from a page when they know the syntax of the page. To put it differently, the traditional techniques can only gather attribute values when the syntax of a page has been previously determined and stored. Accordingly, traditional techniques are generally incapable of gathering information from redesigned and restructured web pages or from new pages because they lack syntax information about those pages. The traditional techniques must first expend effort and resources to determine and store information about syntax before gathering attribute values. (Specification at page 2, lines 4-13.)

This invention is directed to extracting of data records from structured text, such as a web page or any text-containing file, without prior knowledge of the structure of the text. The invention deduces the structure of the text by using information about the attributes and knowledge of candidate structures. (Specification at page 3, lines 9-15.) The claims recite how this is accomplished; in other words, what is claimed is not that data records are extracted without prior knowledge of the structure, but how that is effected.

Independent Claim 43

Independent claim 43 and claims 44 and 45 dependent therefrom are directed to a method (Fig. 2) for extracting records from a structured text in a computer system (100) (specification at page 3, lines

9-10). Claim 43 recites identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields (specification at page 5, lines 23-25; page 7, lines 7-23; blocks 206 and 208 in Fig. 2); identifying a region of interest in the text (specification at page 5, lines 25-27; block 210 of Fig. 2) by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest (specification at page 8, line 4, to page 9, line 7); segmenting the region of interest into record regions that each contain data for a single record (specification at page 5, lines 27-28; block 212 of Fig. 2) by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record field and such that different regions have similar numbers of field values for each record field, selecting one that measures best, applying it to produce record regions, extracting field values from record regions by identifying most likely locations of field values for each record field in each record region (specification at page 9, line 8, to page 10, line 7); and outputting records composed of extracted field values for record fields (specification at page 10, line 8, to page 11, line 10; blocks 214 and 216 of Fig. 2).

Independent Claim 46

Independent claim 46 is directed to an apparatus (Fig. 1) for extracting data from a file, comprising a computer (100) and a computer program (118), performed by the computer (specification at page 3, lines 9-10). Claim 46 recites identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields (specification at page 5, lines 23-25; page 7, lines 7-23; blocks 206 and 208 in Fig. 2); identifying a region

of interest in the text (specification at page 5, lines 25-27; block 210 of Fig. 2) by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest (specification at page 8, line 4, to page 9, line 7); segmenting the region of interest into record regions that each contain data for a single record (specification at page 5, lines 27-28; block 212 of Fig. 2) by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record field and such that different regions have similar numbers of field values for each record field, selecting one that measures best, applying it to produce record regions (specification at page 9, line 8, to page 10, line 7); extracting field values from record regions by identifying most likely locations of field values for each record field in each record region (specification at page 10, lines 8-28; block 214 of Fig. 2); and outputting records composed of extracted field values for record fields (specification at page 11, lines 1-10; block 216 of Fig. 2).

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Rejection of claims 43-46 under 35 U.S.C §102(e) over U.S. patent no. 6,424,980 (Iizuka, et al.).

ARGUMENTS

The disclosure of Iizuka, et al.

Iizuka, et al. represent the prior art referenced by applicants in their "Description of Related Art". On page 1, line 20, to page 2, line 7 of the specification, applicants state:

> Traditional techniques for solving this information gathering problem are typically based on knowledge of the structure used to arrange data within each specific website. (The structure used to arrange the data within a page is commonly referred to as the syntax of the page.) These techniques require prior determination of the syntax of each page and storage of syntax information about each page in a data storage device, such as a database.
>
> When gathering information about a subject from a particular page, the traditional techniques identify the attributes of the subject by comparing the structure of the page with the stored structure information. When there is a match, the traditional technique returns the attribute value to the user.
>
> These traditional techniques are limited because they can only gather attribute values from a page when they know the syntax of a page. To put it differently, the traditional techniques can only gather attribute values when the syntax of a page has been previously determined and stored.

Correspondingly, Iizuka, et al. state:

> The apparatus has a HTML document storing unit for storing meta data about HTML documents. That meta data includes the locations, document structures, presentation locations, presentation styles, etc., of the HTML documents

> for each HTML document.... The document structure data of the HTML documents specifies the structures of partial structure such as tables, lists and clauses contained in the HTML documents and is used to map element data in the table and lists to items to be extracted.
>
> (Col. 11, line 63, to col. 12, line 5)
>
> In the preparatory phase, a managing person prepares meta data about HTML documents through the HTML document meta data manager before starting the execution phase.
>
> (Col. 14, lines 30-32).

In other words, Iizuka, et al. require the syntax of documents that are to be searched to be known and stored before a search can be conducted.

The rejection of claims 43-46

The fundamental difference between applicants' claimed invention and the disclosure of Iizuka, et al. is that Iizuka, et al. require the syntax of documents that are to be searched to be known and stored before a search can be conducted, whereas applicants do not.

The Examiner asserted that this argument is not persuasive because applicants' claims do not recite that data records within a file are identified without using prior knowledge of the structure (e.g., syntax) of the file. This assertion misses the point of applicants' argument. Applicants' argument explains why a teaching of how data records within a document may be identified without knowledge of the structure of a file is not found in Iizuka, et al.: since Iizuka, et al. know the structure a-priori from pre-stored meta-data (see, Iizuka, et al. Figs. 12 and 13, and col. 14, lines 17-21), they do not need to identify it, and consequently they do not teach how it may be identified. In contrast, and as was pointed out above, applicants' claims recite how this identification (and consequent record extraction) is done. Applicants are relying on the functionality – the

particular steps that are recited in the claims – to distinguish their invention from Iizuka, et al. Iizuka, et al. do not disclose, teach, or suggest that functionality.

Inter alia, applicants' claims recite "identifying potential locations of values of record fields in [a structured] text by identifying locations in the text of items in lists of known potential values for record fields." The Examiner asserted that "Iizuka discloses identifying potential locations of values of record fields in the text in Figure 8 at reference signs S200." The Examiner is mistaken. This step of Fig. 8 refers to determining the addresses of documents that are to be searched, in an HTML document table that stores the locations of HTML documents -- see col. 14, lines 15-17 and 47-51 of Iizuka, et al. In contrast, the claim language refers to identifying locations of known potential values for record fields, within a document (text) that is being searched.

The Examiner dismissed this argument by stating that "the examiner equates 'list of known potential values for record fields' with 'HTML document table.'" But applicants' argument cannot be dismissed so easily. The issue is what is being identified. Iizuka, et al. identify addresses of documents that are to be searched, whereas the claims search a text to identify therein potential values (i.e., "items in lists of known potential values") for record fields. The things that are being identified in Iizuka, et al. and in applicants' claims are unmistakably different.

Applicants' claims further recite "identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest." The Examiner asserted that "Iizuka discloses identifying a region of interest in the text by applying candidate region partitions and segmenting the region of interest into record regions that contain data for a single record," and

pointed to Iizuka, et al.'s description of Ashish and Knoblock's technique at col. 2, lines 45-65, as supporting this assertion. The Examiner is again mistaken. This technique identifies the regions (the internal structure) of a text (document). But it does not identify a region of interest among the regions of the text, as required by applicants' claims.

Undaunted, the Examiner dismissed this argument by asserting that "Iizuka discloses 'This technique considers a portion in HTML document as meaningful information' (column 2, line 50, emphasis added). The examiner equates 'a region of interest' with 'portion in HTML document as meaningful information.'" The Examiner's assertion misses the mark. The statement in Iizuka, et al. that "this technique considers a portion in HTML document as meaningful information" merely means that portions of an HTML document are not meaningless -- in other words, that portions, and not only the document as a whole, have meaning. Thus, this statement provides a rationale for why someone would want to determine what portions (regions) a text has. But, significantly, it does not teach identifying a region of interest among those regions, as required by the claims. Nor does it teach identifying the region by the applying, evaluating, and selecting that are recited in the claims. Nor does it teach segmenting the region of interest (as opposed to the text as a whole), as required by the claims. Nor does it effect segmentation by the applying, evaluating, selecting, applying, and extracting that are recited in the claims.

It should therefore by evident that, contrary to the Examiner's assertion, Iizuka, et al. do not disclose, teach, or suggest identifying a region of interest in the text as that identifying is recited in the claims. Nor do they disclose the recited segmenting.

CONCLUSION

For all of the reasons given above, applicants respectfully assert that the Section 102(e) rejection of their appealed claims over Iizuka, et al. is not well founded. Applicants therefore respectfully request that the rejection of the appealed claims be reversed.

Respectfully submitted,

Eric T. Bax
Charles C. Fowlkes
Louis Cisnero, Jr.

By _____
David Volejnicek
Corporate Counsel
Reg. No. 29355
303-538-4154

Date: 04 April 2008

Avaya Inc.
Docket Administrator
307 Middletown-Lincroft Road
Room 1N-391
Lincroft, NJ 07738

The claims on appeal:

43. A method for extracting records from a structured text in a computer system, comprising:

identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields,

identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest,

segmenting the region of interest into record regions that each contain data for a single record by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record field and such that different regions have similar numbers of field values for each record field, selecting one that measures best, applying it to produce record regions, extracting field values from record regions by identifying most likely locations of field values for each record field in each record region, and

outputting records composed of extracted field values for record fields.

44. The method of claim 43, with the addition of:

identifying potential locations of values of record fields in the text by identifying locations in the text of patterns of potential values for record fields.

45. The method of claim 43, with the addition of:

identifying potential locations of values of record fields in the text by identifying locations in the text of numbers in ranges that are potential values for record fields.

46. An apparatus for extracting data from a file, comprising a computer and a computer program, performed by the computer, for:

identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields,

identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest,

segmenting the region of interest into record regions that each contain data for a single record by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record field and such that different regions have similar numbers of field values for each record field, selecting one that measures best, applying it to produce record regions,

extracting field values from record regions by identifying most likely locations of field values for each record field in each record region, and

outputting records composed of extracted field values for record fields.